

<b>REPORT DOCUMENTATION PAGE</b>				Form Approved OMB No. 0704-0188	
<small>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</small>					
<b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.</b>					
<b>1. REPORT DATE (DD-MM-YYYY)</b> 28-02-2010		<b>2. REPORT TYPE</b> Final report		<b>3. DATES COVERED (From - To)</b> 01-06-07 to 30-11-09	
<b>4. TITLE AND SUBTITLE</b> A PDE-Constrained Optimization Approach to Uncertainty Quantification in Inverse Problems, with Applications to Inverse Scattering				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b> FA9550-07-1-0480	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Ghattas, Omar and Biros, George				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Institute for Computational Engineering and Sciences The University of Texas at Austin 1 University Station, C0200 Austin, TX 78712				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Air Force Office of Scientific Research				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b> AFOSR	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b> AFRL-SR-AR-TR-10-0090	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> Approved for public release					
<b>13. SUPPLEMENTARY NOTES</b>				20100316236	
<b>14. ABSTRACT</b> This project addressed the statistical inverse problem of reconstruction of an uncertain shape of a scatterer or properties of a medium from noisy observations of scattered wavefields. The Bayesian solution of this inverse problem yields a posterior pdf, requiring the solution of the forward wave equation to evaluate the probability of any point in parameter space. The standard approach is to sample this pdf via an MCMC method and then compute statistics of the samples. However, standard MCMC methods view the underlying parameter-to-observable map as a black box, and thus do not exploit its structure, becoming prohibitive for high dimensional parameter spaces and expensive simulations. A preconditioned Langevin-accelerated MCMC method for sampling high-dimensional PDE-based probability densities was developed. The preconditioner exploits local Hessian (of the negative log posterior) information to greatly speed up sampling, leading to a stochastic version of Newton's method. Fast Hessian approximations were developed for several inverse scattering problems. Applications to model inverse medium scattering problems indicated several orders of magnitude improvement over a reference black-box MCMC method.					
<b>15. SUBJECT TERMS</b> Statistical inverse problem; inverse shape and medium scattering; Bayesian inference; Langevin dynamics					
<b>16. SECURITY CLASSIFICATION OF:</b> a. REPORT    b. ABSTRACT    c. THIS PAGE U            U            U			<b>17. LIMITATION OF ABSTRACT</b> UU	<b>18. NUMBER OF PAGES</b> 15	<b>19a. NAME OF RESPONSIBLE PERSON</b> Omar Ghattas <b>19b. TELEPHONE NUMBER (Include area code)</b> 512-232-4304

**FINAL REPORT:**  
**A PDE-Constrained Optimization Approach to**  
**Uncertainty Quantification in Inverse Problems with**  
**Applications to Inverse Scattering**

AFOSR Grant FA9550-07-1-0480

Omar Ghattas (PI), Lucas Wilcox (research associate), Tan Bui-Thanh (postdoc),  
James Martin (Ph.D. student)  
Institute for Computational Engineering and Sciences  
The University of Texas at Austin

George Biros (Co-PI), Stephanie Chaillet (postdoc)  
College of Computing  
Georgia Institute of Technology

**Abstract**

This project addressed the statistical inverse problem of reconstruction of an uncertain shape of a scatterer or properties of a medium from noisy observations of scattered wavefields. The Bayesian solution of this inverse problem yields a *posterior pdf*, requiring the solution of the forward wave equation to evaluate the probability of any point in parameter space. The standard approach is to sample this pdf via an MCMC method and then compute statistics of the samples. However, standard MCMC methods view the underlying parameter-to-observable map as a black box, and thus do not exploit its structure, hence becoming prohibitive for high dimensional parameter spaces and expensive simulations.

A Langevin-accelerated MCMC method for sampling high-dimensional PDE-based probability densities was developed. The method builds on previous work in Langevin dynamics, which uses gradient information to guide the sampling in useful directions, improving convergence rates. The Langevin idea was extended to exploit local Hessian (of the negative log posterior) information, leading to a stochastic version of Newton's method. Fast Hessian approximations were developed for several inverse scattering problems. Applications to model inverse medium scattering problems indicated several orders of magnitude improvement over a reference black-box MCMC method.

**1. Approach**

The overall goal of this project has been to create systematic, rigorous, and scalable algorithms for quantifying uncertainties in inverse wave scattering problems. These uncertainties reflect our incomplete knowledge of the medium in which the waves propagate (inverse medium scattering problem) or the shape of a scatterer (inverse shape scattering problem). The problem of inferring an uncertain medium or

shape from observations of scattered wavefields is fundamentally a *statistical inverse problem*. Our lack of knowledge results from noisy measurements, sparse observers, uncertain forward models, and uncertain prior model parameter information. Uncertainty in the reconstructed model parameters is a fundamental feature of ill-posed inverse problems.

The deterministic approach to the inverse scattering problem, which amounts to minimizing a regularized data misfit function, is incapable of accounting for uncertainties in the solution of the inverse problem. *Bayesian inference* provides a systematic framework for incorporating uncertainties in observations, forward models, and prior knowledge to quantify uncertainties in the model parameters [29, 32]. Suppose the relationship between output observables  $y$  (such as waveforms at sensor locations) and uncertain model parameters  $p$  (such as those describing a wave speed of a heterogeneous medium or shape of a scatterer) is denoted by  $y = f(p, e)$ , where  $e$  represents noise due to measurement and/or modeling errors. In other words, given the model parameters  $p$  and noise  $e$ , the function  $f(p, e)$  solves the forward (acoustic, elastic, or electromagnetic wave propagation) problem to yield  $y$ , the predicted outputs at the measurement locations (and time instants). Suppose also that we have the prior probability density  $\pi_{\text{prior}}(p)$ , which encodes the confidence we have in prior information on the unknown model parameters (i.e. independent of present observations), and the likelihood function  $\pi(y_{\text{obs}}|p)$ , which describes the conditional probability that the model parameters  $p$  gave rise to the actual measurements  $y_{\text{obs}}$ . Then Bayes' theorem of inverse problems expresses the posterior probability density of the model parameters,  $\pi_{\text{post}}$ , given the data  $y_{\text{obs}}$ , as the conditional probability

$$\pi_{\text{post}}(p) \stackrel{\text{def}}{=} \pi(p|y_{\text{obs}}) \propto \pi_{\text{prior}}(p) \pi(y_{\text{obs}}|p). \quad (1)$$

Expression (1) provides the statistical solution of the inverse problem as a probability density for the model parameters  $p$ .

As a specific example, suppose the noise is additive and is modeled as Gaussian with zero mean and a covariance matrix  $C_{\text{noise}}$ , and suppose the prior density of the model parameters is represented as Gaussian with  $p_{\text{prior}}$  as the mean and  $C_{\text{prior}}$  as the covariance matrix, then the posterior probability density of the model parameters is given explicitly (within a normalizing constant) by

$$\pi_{\text{post}}(p) \propto \exp \left[ -\frac{1}{2} (f(p) - y_{\text{obs}})^T C_{\text{noise}}^{-1} (f(p) - y_{\text{obs}}) - \frac{1}{2} (p - p_{\text{prior}})^T C_{\text{prior}}^{-1} (p - p_{\text{prior}}) \right]. \quad (2)$$

This latter expression shows that even when the prior, measurement, and modeling uncertainties are Gaussian, the posterior density of the model parameters is generally not Gaussian, due to the nonlinearity of the parameter-to-observable map,  $f(p)$ . However, this expression exposes a significant connection between statistical and deterministic inversion. Suppose we wish to find the value of the most likely model parameters, by maximizing the posterior density (2). This is equivalent to minimizing the negative argument of the exponential function—which is precisely the misfit function that is minimized by deterministic inverse methods, provided we interpret

the prior as a regularization and weigh the data misfit by the inverse noise covariance. Moreover, it is straightforward to show that the inverse of the Hessian matrix of the deterministic misfit function approximates the covariance matrix of the posterior density (the equivalence is exact when  $f(p)$  is linear). This connection between the Hessian operator of the deterministic inverse problem and the inverse covariance matrix of the statistical inverse problem is crucial, and we believe is a key (and unexploited) idea in overcoming the curse of dimensionality associated with uncertainty quantification in inverse problems.

While it is easy to write down expression such as (1) or (2) for the posterior probability density, making use of these expressions poses a challenge, because the posterior probability density is a surface in high dimensions (equal to the number of model parameters  $p$ ), and because the solution of the forward problem (i.e., computing  $f(p)$  given  $p$ ) is required to evaluate the probability of any point in parameter space. Straightforward grid-based sampling is out of the question for anything other than a few parameters and cheap forward simulations. Special sampling techniques, such as Markov chain Monte Carlo (MCMC) methods, have been developed to generate sample ensembles that typically require many fewer points than grid-based sampling [21, 29, 32, 33].

In particular, Metropolis-Hastings (M-H) methods employ a given probability density  $q(p_k, y)$  at each sample point in parameter space  $p_k$  to generate a proposed sample point  $y$ . Once generated, the M-H criterion chooses to either accept or reject the proposed sample point depending on its probability relative to the probability of the current point, and repeats from the new point, thereby generating a chain of samples from the posterior density  $\pi_{\text{post}}(p)$ . For example, a popular choice for the proposal density is the isotropic Gaussian  $q(p_k, y) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}(p_k - y)^2]$ ; the resulting method is known as Random Walk Metropolis. Though easy to sample from, this choice of proposal function is not tailored to the structure of the underlying posterior probability, and is therefore typically slow to converge. MCMC methods such as Random Walk Metropolis become prohibitive as the complexity of the forward simulation and the dimension of the parameter space increase. When the model parameters represent a (suitably-discretized) field (such as scatterer shape or medium wave speed), and when the forward PDE requires hours to solve on a parallel computer (such as mid-to-high frequency 3D wave propagation), the MCMC framework becomes *completely intractable*.

The central difficulty in scaling up conventional MCMC for large-scale forward simulations and high-dimensional parameter spaces is that this is a purely *black-box* approach, i.e. it does not exploit the structure of the parameter-to-observable map  $f(p)$ . Twenty years of advances in algorithms for *deterministic* large-scale PDE-constrained optimization have taught us that making maximal use of derivative information can greatly speed up the search process for extremum points (e.g. [7, 11, 12, 13, 27, 28]). In this project we have aimed to overcome the intractability of conventional methods for statistical inverse scattering problems by developing scalable algorithms that exploit the structure of inverse wave propagation parameter-to-observable operators

$f(p)$ . We have focused on developing *preconditioned Langevin methods* that exploit this structure to greatly speed up sampling (Section 2). A crucial aspect of the success of preconditioned Langevin methods are fast algorithms for approximating the inverse Hessian for inverse wave scattering problems (Section 3).

## 2. Fast Hessian-preconditioned Langevin samplers

We have developed fast sampling methods that build on—and significantly extend—ideas from Langevin dynamics, which use gradient information to accelerate sampling of a target density, e.g. [10, 31]. The Langevin equation is a stochastic differential equation,

$$d\mathbf{P}_t = \mathbf{A} \nabla \log \pi_{\text{post}} dt + \sqrt{2} \mathbf{A}^{1/2} d\mathbf{W}_t, \quad (3)$$

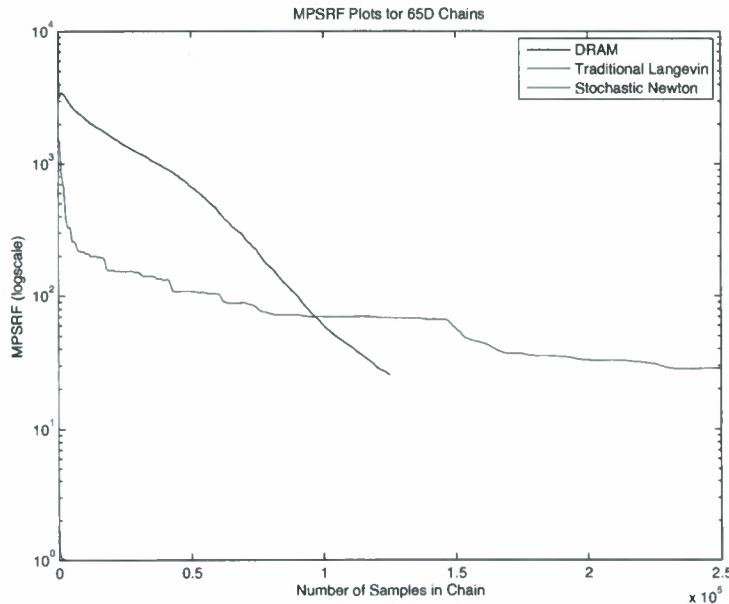
with  $\pi_{\text{post}}(p)$  as an invariant density. Here,  $\mathbf{W}_t$  is the i.i.d. vector of standard Brownian motions. Preconditioning by a symmetric positive definite operator  $\mathbf{A}$  preserves the invariance of the density. In practice, we discretize in time with timestep  $\Delta t$ , yielding (e.g. for explicit Euler) the update

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{A} \nabla \log \pi_{\text{post}} \Delta t + \sqrt{2\Delta t} \mathbf{A}^{1/2} \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (4)$$

where  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  is the i.i.d. standard normal density. Discretization in time can add bias, so we use the Langevin steps as proposals for a Metropolis-Hastings MCMC method. The form (4) shows immediately the connection with deterministic optimization methods: the gradient term  $\nabla \log \pi_{\text{post}}$  is a steepest ascent direction for the posterior density. In its absence (and in the absence of preconditioning, i.e.  $\mathbf{A} = \mathbf{I}$ ) we recover a Gaussian random walk from the last term in (3). The addition of the gradient term drives the sampling in (the locally steepest) direction of higher probability. While bringing in gradient information to accelerate sampling is attractive, we know that steepest descent is a poor choice for search directions in large-scale optimization (particularly for anisotropic pdfs), and we seek to improve on it.

Taking the preconditioner  $\mathbf{A}$  as the inverse of the Hessian of  $\log \pi_{\text{post}}$ , we obtain the stochastic equivalent of Newton’s method. In the common case of Gaussian additive noise and prior, the (negative) log of the posterior density is simply the “regularized” misfit function (the sum of the data misfit and prior/regularization term) that deterministic inverse methods seek to minimize. Thus, similar to Newton’s locally-quadratic approximation of the objective, the Hessian-preconditioned Langevin step makes a locally-Gaussian approximation of  $\pi_{\text{post}}$ . This endows the sampling process with information on the curvature of the posterior density surface, which is crucial in high dimensions. This results in a need for substantially fewer sampling points relative to a black-box MCMC method, just as deterministic Newton requires substantially fewer iterations to find the optimum compared to a derivative-free optimization method.

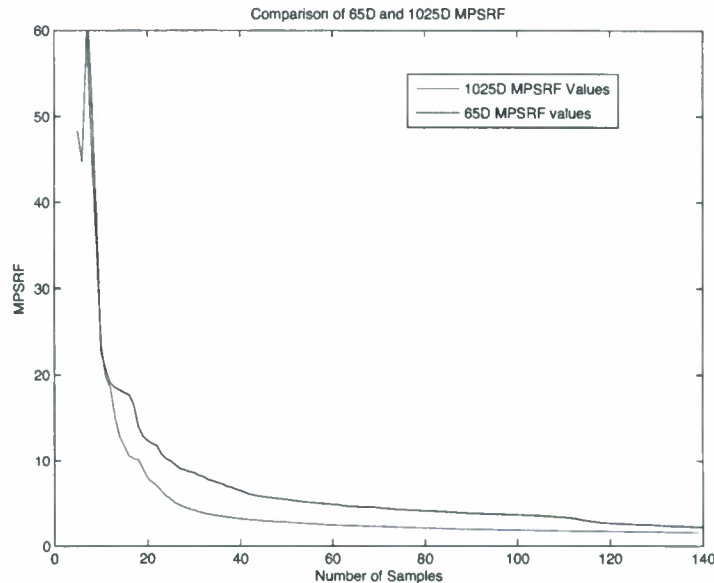
Moreover, it can be shown [20] that in the limiting case when the posterior density  $\pi_{\text{post}}$  is in fact Gaussian (e.g. when the inverse problem is linear and the noise is additive and Gaussian), this so-called *stochastic Newton* method not only samples the target density at long times, but accurately samples from  $\pi_{\text{post}}$  at *every time step*. This



**Figure 1:** Comparison of number of points taken for sampling posterior density for a 1D inverse scattering problem to identify the distribution of the elastic modulus of a heterogeneous medium for a 65-dimensional discretization of the medium. DRAM (black), unpreconditioned Langevin (blue), and Stochastic Newton (red) sampling methods are compared. Convergence indicator is multivariate potential scale reduction factor (MPSRF [18]), for which a value of unity indicates convergence. Stochastic Newton requires over *three orders of magnitude* fewer sampling points.

means that the Metropolis-Hastings criterion will accept *all* of the proposed sample points, and that a minimum number of points are needed to accurately sample from the given distribution. For densities that are not Gaussian, stochastic Newton will still provide a substantial speedup over a conventional random walk, since a Gaussian approximation (based on a local quadratic approximation of  $\log \pi_{\text{post}}$ , or equivalently a linearized approximation of the inverse problem) will generally yield more useful information on the behavior of  $\pi_{\text{post}}$  than a standard normal density approximation (or other heuristic) will.

We have developed an implementation of the stochastic Newton method, and applied it to solve nonlinear inverse medium and shape scattering problems in one and two dimensions [20]. For example, for a 1D inverse medium problem in which the uncertain elastic modulus of the medium is discretized into 64 piecewise linear finite elements, Figure 1 indicates just  $\mathcal{O}(10^2)$  samples are necessary to adequately sample the (non-Gaussian) posterior density, while a reference publicly-available (non-derivative) MCMC method (Delayed Rejection Adaptive Metropolis (DRAM)) [25]) is nowhere near converged after even  $\mathcal{O}(10^5)$  samples. The performance of unpreconditioned Langevin MCMC is similar to that of DRAM, indicating the crucial role of the Newton direction vs. steepest descent. Moreover, because the (inverse) Hessian captures the (local) covariance structure of the posterior density, this orders-of-magnitude speedup is expected to become even larger as the parameter dimension increases. Furthermore, the stochastic Newton method is able to sample a poste-

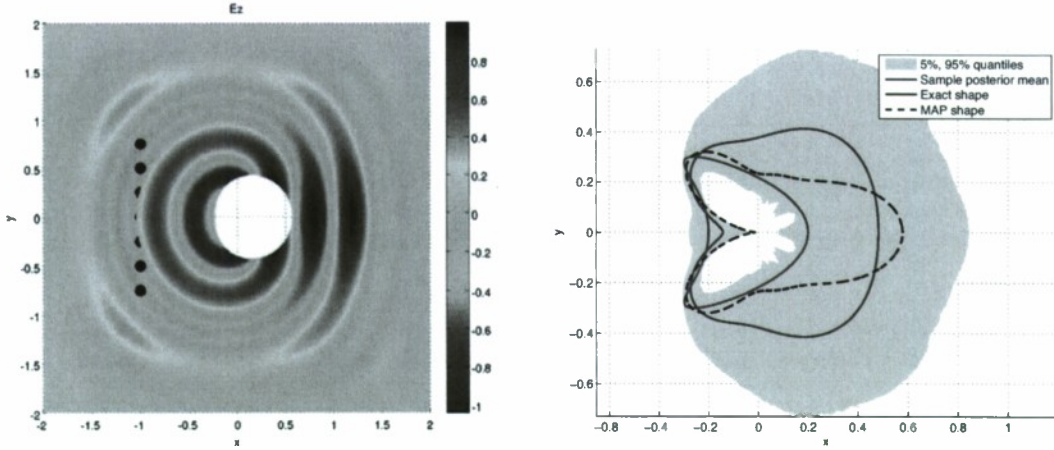


**Figure 2:** Convergence of stochastic Newton for 1025-dimensional problem compared with 65-dimensional problem (1025-dimension results based on fast low-rank implementation), showing similar rates of convergence. Figure implies mesh-independence (i.e. dimension-independence) of stochastic Newton method for this problem.

rior pdf stemming from a 1025-dimensional problem (in which the wave propagation medium is discretized into 1024 finite elements). Figure 2 compares the convergence of the stochastic Newton method for the 65-dimensional and 1025-dimensional inverse medium problems. As can be seen in the figure, the convergence behavior for the two problems is similar; *in other words, the convergence is independent of problem size*. This behavior is similar to the well-known mesh-independence property of the deterministic Newton method.

The stochastic Newton method has also been used to solve the electromagnetic statistical inverse problem of inferring uncertainty in the shape of a scatterer from the scattered wavefield. Based on the forward code from [26], we have built a 2D discontinuous Galerkin spectral element time-domain Maxwell solver, truncated with PMLs and enhanced with adjoint-based shape gradient/Hessian capability, and numerically integrated in time with a fourth-order, five-stage explicit Runge Kutta scheme. Figure 3 compares the uncertain shape reconstructed from scattered noisy waveforms, with the “exact” shape. As can be seen from the figure, the statistical solution to the inverse problem (implied by grey shading) goes well beyond the deterministic solution (see blue<sup>1</sup> shape), by conveying information about the confidence we have in the inferred shape—taking into account any prior knowledge on the shape and noise in the data and model. In this case, the front of the scatterer is identified with notably less

<sup>1</sup>We are speaking only loosely when we equate the mean shape with the deterministic solution. If the noise is additive and Gaussian and the parameter-to-observable map is linear, then the two are equivalent; otherwise, they will differ, and the differences will grow as the map becomes more nonlinear.



**Figure 3:** Uncertain shape reconstructed from noisy scattered EM observations. Discontinuous Galerkin discretization with 3rd order spectral elements in space and 4th-order, 5-stage explicit Runge Kutta integration in time. Prior favors shape with small boundary. Shape is parametrized by 6 cosine modes,  $r(\theta) = \sum_{i=0}^5 a_i \cos(\theta)$ . Computational domain  $\Omega = \{(x, y) : -1 \leq x, y \leq 1\}$ . PML domain  $\Omega_{PML} = \{(x, y) : 1 \leq |x|, |y| \leq 2\}$ . Kite shape to generate synthetic observations given by  $x = 0.2[\cos(\theta) + 0.65(\cos(2\theta) - 1)]$ ,  $y = 0.3 \sin(\theta)$ . Incident wave  $E_z^I = \cos(8(t-x))$ ,  $H_x = 0$ ,  $H_y = 0$  from left. 31 observation points:  $x = -0.9, y = \text{linspace}\{-0.9, 0.9, 31\}$ .  $E_z, H_x, H_y$  are observed in  $0 \leq t \leq \pi$  at all time steps with 5% Gaussian noise.  $\Delta t = 10^{-3}$ , resulting in 3324 time steps. Mesh size  $h_{\min} = 0.05$ , resulting in  $\sim 135,000$  DOF. *Left:* Snapshot of electric field. Plane wave incident from left, receivers located at black dots, scatterer shown in white. *Right:* Comparison of “exact” shape (red) with mean of reconstructed shape (blue) and maximum a posteriori shape (black), superposed on gray region indicating 5% and 95% quantiles of shape uncertainty. Note that uncertainty in reconstruction is greatest behind the scatterer.

uncertainty than the back, and in general the recovered shape has large uncertainty, stemming primarily from the limited observations. Despite the statistical inverse problem’s having “only” six parameter, the size of the forward problem (135,000 spatial unknowns and over 3000 time steps) make forward solution so expensive that the problem cannot be solved at all by conventional MCMC.

The critical role that inverse Hessian preconditioning plays is that it brings to the statistical sampling process the power of Newton-type deterministic optimization methods—which for a large class of PDE-based inverse problems can deliver solutions at the cost of a constant number of forward solves, independent of problem size, e.g. [4, 5, 6, 7, 9, 16, 17, 19, 23]. However, stochastic Newton carries the added cost of having to compute the deterministic Newton step (i.e. solve linear systems having the Hessian as coefficient matrix) as well as find a square root of the inverse of the Hessian at each sample point, as seen in Equation (4). Thus the cost of a stochastic Newton sample point can be significantly more expensive than that of a non-derivative MCMC method. Explicitly constructing the (formally dense) Hessian operator and factoring it is completely out of the question for large-scale problems, since this would require a number of forward solves equal to the number of uncertain model parameters. However, the key to the success of Newton methods for deterministic inverse problems is to recognize that for ill-posed inverse problems, the Hessian

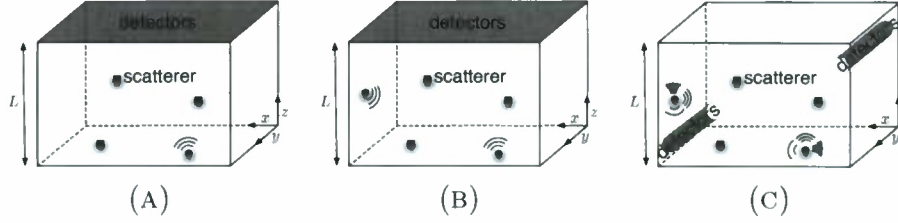
(or its Gauss-Newton approximation) is usually the sum of a compact operator (the data misfit part, since the data typically provide limited information on the model parameter fields) and a differential part (since the regularization/prior is typically smoothing, and its inverse is seen by the Hessian). Thus, when preconditioned by the prior, the Hessian often behaves like a compact perturbation of the identity, and fast approximations (e.g. using low rank [8] or multilevel approximations [1, 5], or spectral representations [14]) can be combined with Krylov methods to find the action of the inverse Hessian on a vector in a mesh-independent (and often small) number of forward solves. Moreover, these ideas can be extended to rapidly find a square root of the inverse Hessian [20], which is needed to draw samples from the local Gaussian approximation of the posterior.

Incorporating Hessian information in Langevin dynamics-based sampling as above permits explicit separation of (1) the data misfit contribution to the posterior, which typically provides “sparse” information, i.e. information on a limited number of directions in parameter space (a reflection of the ill-posed nature of the inverse problem), meaning that the expensive forward simulations—needed to relate model parameters to observables—can be limited to just these directions, and (2) the prior contribution to the posterior, which often provides “dense” information in parameter space, but this information is independent of the forward model, and thus is cheap to evaluate. The typically small and bounded dimension of the range space of the data misfit component of the Hessian thus plays a critical role in dimensionality reduction, and we believe is key to overcoming the curse of dimensionality for PDE-based statistical inverse problems. Clearly an important issue is the construction of fast low rank estimates of the data misfit portion of the Hessian for different classes of inverse operators, and in particular for inverse wave propagation. This is discussed in Section 3 for time harmonic inverse medium scattering.

Finally, all of the important computational kernels of stochastic Newton resemble those of large-scale deterministic Newton-based inverse solvers (notably the solution of forward and adjoint state problems, and the combination of their solutions to form gradients and actions of Hessians on vectors). This permits the exploitation of highly scalable parallel data structures, algorithms, and implementations that have been developed for deterministic inverse problems, and have been used to solve the PDE-based inverse problems with up to  $\mathcal{O}(10^8)$  inversion parameters on multi-thousand processor systems [4, 5, 6, 7, 23].

### 3. Fast Hessian approximations for time-harmonic inverse medium scattering

As discussed above, the ability to cheaply approximate the Hessian of the data misfit functional is critical to the success of the stochastic Newton method. Here we describe **FaIMS** (Fast Inverse Medium Solver), a novel algorithm for the construction of fast Hessian approximations for the low-frequency Helmholtz inverse medium problem with broadband and multi-point illuminations. Inverse medium problems are encountered in acoustic, elastic, and electromagnetic wave propagation. We use a Lippmann-Schwinger formulation, which we discretize using a quadrature method.



**Figure 4:** FaIMS is a fast method for the solution of a Born approximation inverse medium problem in the low frequency regime. The medium perturbation is represented by a set of point scatterers (red squares) in a 3-D domain  $H$ . The data consist of measurements of scattered fields due to several different incident fields. For example, in subfigure (A) we depict a single-source, single-frequency illumination, in (B) we depict a multiple-source, multiple-frequency illumination in which we collect data for each different frequency and in (C) we depict the case of multiple-source multiple-frequency illumination. The detectors can be located in arbitrary positions (C).

We consider small perturbations of the background medium.

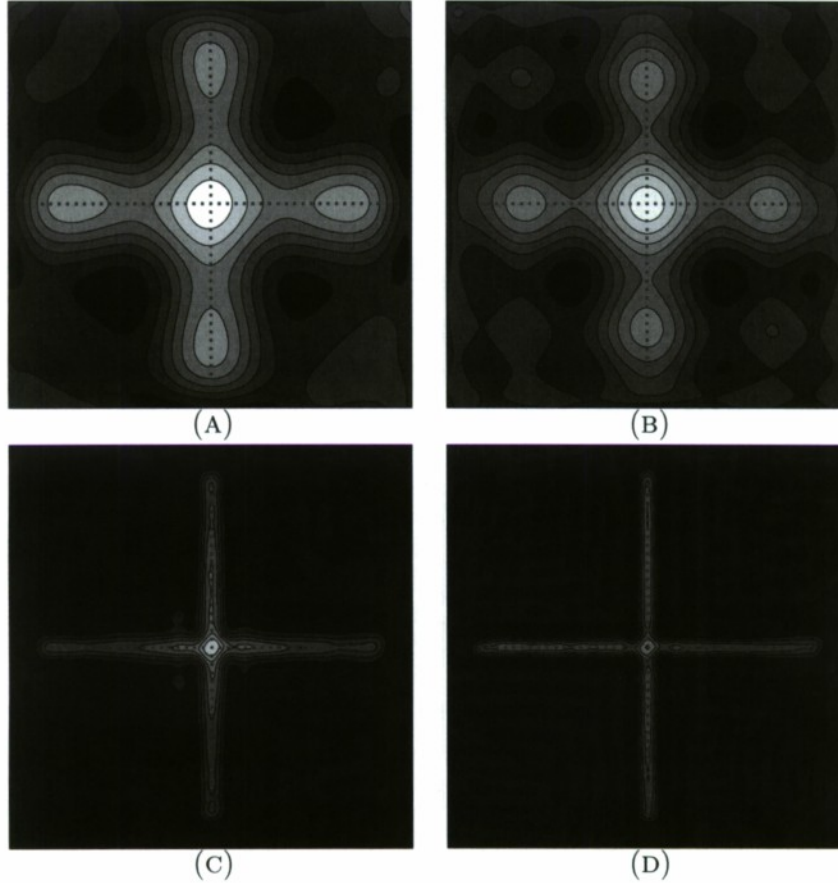
If  $N_\omega$  is the number of excitation frequencies,  $N_s$  the number of different source locations for the point illuminations,  $N_d$  the number of detectors, and  $N$  the scatterer discretization size, a dense singular value decomposition for the overall input-output map (roughly speaking, the square root of the Hessian) will require  $\mathcal{O}([\min(N_s N_\omega N_d, N)]^2 \times \max(N_s N_\omega N_d, N))$  operations, without accounting for the costs of solving the forward problem. We have developed a fast SVD approach that brings the cost down to  $\mathcal{O}(N_s N_\omega N_d N)$  thus, providing orders of magnitude improvements over a black-box dense SVD. The method is also more robust and readily parallelizable when compared to Krylov-based SVD approaches. FaIMS builds on our previous work on fast Hessian [1, 2, 3, 15] approximations and can be combined with multigrid methods such as the ones developed in [1, 5]. The work is described in detail in [22]. Here, we give some details on the problem statement and some representative results.

Given  $N_s$  point sources that generate  $N_s N_\omega$  incident fields (spherical waves)  $\{u(\mathbf{r}; s, \omega)\}_{s=1, \omega=1}^{N_s, N_\omega}$ , at  $N_\omega$  different frequencies, we measure the scattered field  $\phi(\mathbf{r}_d; s, \omega)$  at  $N_d$  detector locations  $\{\mathbf{r}_d\}_{d=1}^{N_d}$ . We seek to recover the medium perturbation  $\eta(\mathbf{r})$ , by solving

$$\phi(\mathbf{r}_d; s, \omega) = \int_H G(\mathbf{r}_d, \mathbf{r}; \omega) \eta(\mathbf{r}) u(\mathbf{r}; s, \omega) d\mathbf{r}, \quad \omega = \omega_1, \dots, \omega_{N_\omega}, \quad s = 1, \dots, N_s. \quad (5)$$

This is a Born-approximation Lippmann-Schwinger scattering equation, where  $G(\cdot, \cdot; \omega)$  is the Green's function (in the reference medium) at frequency  $\omega$ ,  $H$  is the support of the medium perturbation  $\eta$ , and  $\mathbf{r}$  is a point in  $H$ . This equation is discretized using  $N$  quadrature points. The problem setup is summarized in Figure 4.

One solution approach would be to use a dense SVD factorization. However, such an approach is prohibitively expensive, as the complexity of a dense SVD is  $[\min(N_s N_\omega N_d, N)]^2 \times \max(N_s N_\omega N_d, N)$  [24]. (For example, if  $N_\omega = 10$ ,  $N_s = 100$ ,  $N_d = 10^2$ , and  $N = 100^3$ , we will need over one month of computation to compute the SVD on a single-core,



**Figure 5:** We consider a scatterer perturbation with a cross-geometry, with a 10% contrast with the background medium and we consider different scatterer sizes, up to ten wavelengths. In this experiment, we verify the accuracy and efficiency of our scheme. We test FaIMS by comparing its reconstruction (subfigure A,C) to the one obtained by solving the output least squares inverse problem using the LSQR algorithm (A) and (B) corresponding to the case in which the cross size is one wavelength. Subfigures (C) and (D) correspond to the case in which the cross size is ten wavelengths. In both cases, we are able to achieve excellent reconstruction accuracy using FaIMS at considerably cheaper cost.

two-Gigaflops/sec machine.) Our goal is to design an algorithm that reconstructs  $\boldsymbol{\eta}$  and scales well with  $N$ ,  $N_s$ ,  $N_\omega$  and  $N_d$ , for the low frequency regime. Our main contribution is the construction of an approximate singular value decomposition that is valid for arbitrary distributions of sources, detectors and frequencies, as long as the detectors are well separated from the support of the scatterer. Our construction is based on the following algorithmic components:

- rank-revealing randomized factorization ideas;
- preprocessing of the incident field  $u$  using SVD to transform the incoming field and data and reduce the dimension of  $N_s$ ;

- and a novel recursive SVD algorithm that can be used to compute the SVD of  $\mathbf{M} = [\mathbf{M}_1 \ \mathbf{M}_2]$  given the SVDs of  $\mathbf{M}_1$  and  $\mathbf{M}_2$ .

Using these components, we propose an approximate SVD factorization whose total work complexity is  $O(N_s N_\omega N_d N)$ . The main idea is to decompose the Hessian operator into  $N_s \times N_\omega$  smaller submatrices of size  $N_d \times N$  ( $1 \leq \omega \leq N_\omega$ ,  $1 \leq s \leq N_s$ ). We compute the SVD of each small matrix by using the randomized SVD proposed in [30]. We apply a low rank approximation whenever possible, leading to a compression of the matrix and a speed up of the computations. Then, we recursively compute the SVD of the Hessian given the SVDs of the smaller submatrices, using a novel method we have devised. The SVD provides a precise characterization of the inverse problem since it allows us to easily apply pseudo-inverse of the Hessian to the data. An example of the accuracy of this approximate SVD is presented in Figure 5. Fast and accurate Hessian approximations such as this play an important role in making the Hessian-preconditioned Langevin method of Section 2 scalable to large problem sizes.

### Acknowledgment/Disclaimer

This work was sponsored (in part) by the Air Force Office of Scientific Research, USAF, under grant number FA9550-07-1-0480. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

### Publications

1. S.S. Adavani and G. Biros, *Fast Algorithms for Source Identification Problems with Elliptic PDE Constraints*, **SIAM Journal on Imaging Sciences**, 2010, in press.
2. S.S. Adavani and G. Biros, *Fast Algorithms for Source Identification Problems with Parabolic PDE Constraints*, **SIAM Journal on Imaging Sciences**, 2010, in press.
3. S.S. Adavani and G. Biros, *Multigrid Algorithms for Inverse Problems with Linear Parabolic PDE Constraints*, **SIAM Journal on Scientific Computing**, 31(1), 369–397, 2008.
4. V. Akcelik, H.P. Flath, O. Ghattas, J. Hill, B. van Bloemen Waanders, and L.C. Wilcox, *Fast Algorithms for Bayesian Uncertainty Quantification in Large-Scale Linear Inverse Problems Based on Low-Rank Partial Hessian Approximations*, **SIAM Journal on Scientific Computing**, submitted, 2009.
5. A. Askan, V. Akcelik, J. Bielak, and O. Ghattas, *Parametric Analysis of a Nonlinear Least Squares Optimization-Based Anelastic Full Waveform Inversion Method*, **Comptes Rendus Mecanique**, submitted.
6. L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders, and K. Willcox (eds), *Large*

- Scale Inverse Problems and Quantification of Uncertainty*, John Wiley & Sons, 2010, to appear.
7. G. Biros and G. Doğan, *A Multilevel Algorithm for Inverse Problems with Elliptic PDE Constraints*, **Inverse Problems**, 24(3), (19pp), 2008.
  8. C. Burstedde and O. Ghattas, *Algorithmic Strategies for Full Waveform Inversion: 1D Experiments*, **Geophysics**, 74(6):WCC37–WCC46, 2009.
  9. C. Burstedde, O. Ghattas, J. Martin, and L.C. Wilcox, *Uncertainty Quantification in Inverse Problems with Stochastic Newton MCMC*, in preparation, 2010.
  10. S. Chaillat and G. Biros, *A Fast Algorithm for Multi-Frequency and Multipoint Inverse Scattering Problems*, in preparation, 2010.
  11. I. Epanomeritakis, V. Akcelik, O. Ghattas, and J. Bielak, *A Newton-CG Method for Large-Scale Three-Dimensional Elastic Full-Waveform Seismic Inversion*, **Inverse Problems**, 24:034015 (26pp), 2008.
  12. L. Wilcox, G. Stadler, C. Burstedde, O. Ghattas, *A High-Order Discontinuous Galerkin Method for Wave Propagation Through Coupled Elastic-Acoustic Media*, Manuscript to be submitted to **Journal of Computational Physics**, 2010.

## References

- [1] S. S. ADAVANI AND G. BIROS, *Multigrid algorithms for inverse problems with linear parabolic PDE constraints*, SIAM Journal on Scientific Computing, 31 (2008), pp. 369–397.
- [2] ———, *Fast algorithms for source identification problems with elliptic PDE constraints*, SIAM Journal on Imaging Sciences, (2010), p. 18. In press.
- [3] ———, *Fast algorithms for source identification problems with parabolic PDE constraints*, SIAM Journal on Imaging Sciences, (2010). In press.
- [4] V. AKÇELIK, J. BIELAK, G. BIROS, I. EPANOMERITAKIS, A. FERNANDEZ, O. GHATTAS, E. J. KIM, J. LOPEZ, D. R. O’HALLARON, T. TU, AND J. URBANIC, *High resolution forward and inverse earthquake modeling on terascale computers*, in Proceedings of SC2003, Phoenix, 2003, ACM/IEEE. Gordon Bell Prize for Special Achievement.
- [5] V. AKÇELIK, G. BIROS, A. DRAGANESCU, O. GHATTAS, J. HILL, AND B. VAN BLOEMAN WAANDERS, *Dynamic data-driven inversion for terascale simulations: Real-time identification of airborne contaminants*, in Proceedings of SC2005, Seattle, 2005.
- [6] V. AKÇELIK, G. BIROS, AND O. GHATTAS, *Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation*, in Proceedings of IEEE/ACM SC2002 Conference, Baltimore, MD, Nov. 2002. SC2002 Best Technical Paper Award.

- [7] V. AKÇELIK, G. BIROS, O. GHATTAS, J. HILL, D. KEYES, AND B. VAN BLOEMAN WAANDERS, *Parallel PDE constrained optimization*, in *Parallel Processing for Scientific Computing*, M. Heroux, P. Raghaven, and H. Simon, eds., SIAM, 2006.
- [8] V. AKÇELIK, H. FLATH, O. GHATTAS, J. HILL, B. VAN BLOEMEN WAANDERS, AND L. WILCOX, *Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations*. Submitted, 2009.
- [9] A. ASKAN, V. AKÇELIK, J. BIELAK, AND O. GHATTAS, *Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures*, *Bulletin of Seismological Society of America*, 97 (2007), pp. 1990–2008.
- [10] A. BESKOS AND A. STUART, *MCMC methods for sampling function space*, in *Plenary Lectures ICIAM 2007*, to appear.
- [11] L. BIEGLER, G. BIROS, O. GHATTAS, M. HEINKENSCHLOSS, D. KEYES, B. MALLICK, Y. MARZOUK, L. TENORIO, B. VAN BLOEMEN WAANDERS, AND K. WILLCOX, eds., *Large Scale Inverse Problems and Quantification of Uncertainty*, John Wiley & Sons, 2010. To appear.
- [12] L. T. BIEGLER, O. GHATTAS, M. HEINKENSCHLOSS, D. KEYES, AND B. VAN BLOEMEN WAANDERS, eds., *Real-Time PDE-Constrained Optimization*, SIAM, 2007.
- [13] L. T. BIEGLER, O. GHATTAS, M. HEINKENSCHLOSS, AND B. VAN BLOEMEN WAANDERS, eds., *Large-Scale PDE-Constrained Optimization*, *Lecture Notes in Computational Science and Engineering*, Vol. 30, Springer-Verlag, Heidelberg, 2003.
- [14] G. BIROS AND G. DOĞAN, *A fast algorithm for an inverse medium problem with multiple measurements*. Submitted for publication, 2008.
- [15] G. BIROS AND G. DOĞAN, *A multilevel algorithm for inverse problems with elliptic PDE constraints*, *Inverse Problems*, 24 (2008), p. 034010 (18pp).
- [16] G. BIROS AND O. GHATTAS, *Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part I: The Krylov–Schur solver*, *SIAM Journal on Scientific Computing*, 27 (2005), pp. 687–713.
- [17] —, *Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part II: The Lagrange–Newton solver and its application to optimal control of steady viscous flows*, *SIAM Journal on Scientific Computing*, 27 (2005), pp. 714–739.
- [18] S. BROOKS AND A. GELMAN, *General methods for monitoring convergence of iterative simulations*, *Journal of Computational and Graphical Statistics*, 7 (1998), pp. 434–456.

- [19] C. BURSTEDDE AND O. GHATTAS, *Algorithmic strategies for full waveform inversion: 1D experiments*, Geophysics, 74 (2009), pp. WCC37–WCC46.
- [20] C. BURSTEDDE, O. GHATTAS, J. MARTIN, AND L. C. WILCOX, *Uncertainty quantification in inverse problems with stochastic Newton MCMC*. Manuscript, 2009.
- [21] D. CALVETTI AND E. SOMERSALO, *Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, Springer, New York, 2007.
- [22] S. CHAILLAT AND G. BIROS, *FaIMS: A fast algorithm for the inverse medium problem with multiple frequencies and multiple sources for the scalar Helmholtz equation*. In preparation, 2010.
- [23] I. EPANOMERITAKIS, V. AKÇELIK, O. GHATTAS, AND J. BIELAK, *A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion*, Inverse Problems, 24 (2008), p. 034015 (26pp).
- [24] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, third ed., 1996.
- [25] H. HAARIO, M. LAINE, A. MIRAVETE, AND E. SAKSMAN, *DRAM: Efficient adaptive MCMC*, Statistics and Computing, 16 (2006), pp. 339–354.
- [26] J. S. HESTHAVEN AND T. WARBURTON, *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*, vol. 54 of Texts in Applied Mathematics, Springer, 2008.
- [27] M. HINZE, R. PINNAU, M. ULBRICH, AND S. ULBRICH, *Optimization with PDE Constraints*, Springer, 2009.
- [28] K. ITO AND K. KUNISCH, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM, 2008.
- [29] J. KAIPIO AND E. SOMERSALO, *Statistical and Computational Inverse Problems*, vol. 160 of Applied Mathematical Sciences, Springer-Verlag, New York, 2005.
- [30] E. LIBERTY, F. WOOLFE, P. MARTINSSON, V. ROKHLIN, AND M. TYGERT, *Randomized algorithms for the low-rank approximation of matrices*, Proceedings of the National Academy of Sciences, 104 (2007), p. 20167.
- [31] A. STUART, P. WIBERG, AND J. VOSS, *Conditional path sampling of SDEs and the Langevin MCMC method*, Communications in Mathematical Sciences, 2 (2004), pp. 685–697.
- [32] A. TARANTOLA, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, PA, 2005.

- [33] L. TIERNEY, *Markov chains for exploring posterior distributions*, The Annals of Statistics, 22 (1994), pp. 1701–1762.